Exercise 1: Text File Import in Databricks

filePath = "dbfs:/databricks-datasets/SPARK\_README.md"

Question 0: How to check file system?

Question 1: How to load a text file?

Question 2: Show me first 10 lines in the file

Question 3: Show me number of lines

Question 4: Show me first line

Question 5: Show me first 10 lines in the file

Question 6: Show me how many times “spark” word contains

Question 7: Split each line into a list of words separated by a space from the dataset

Question 8: **Filter the list of words to exclude common stop words**

Question 9: **Cache the filtered dataset in memory to speed up future actions.**

Question 10: **Transform filtered words into list of (word,1) tuples for WordCount**

Question 11: **Aggregate the (word,1) tuples into (word,count) tuples**

Question 12: **Display the top 10 (word,count) tuples by count**

Question 13: **Create a table from the (word,count) tuples**